

---

# THEORY AND COMPUTING OF GAS PHASE CHEMICAL REACTIONS: FROM EXACT QUANTUM TO APPROXIMATE DYNAMICAL TREATMENTS

A. Laganà

\*

*Dipartimento di Chimica, Università di Perugia,  
06123 Perugia, Italy*

8 February 2006

## Contents

<b>1 Abstract</b>	<b>3</b>
<b>2 Introduction</b>	<b>3</b>
<b>3 The <i>a priori</i> dynamical approach</b>	<b>4</b>
3.1 Electronically adiabatic potential energy surfaces . . . . .	5
3.2 Quantum formalism for few body reactions . . . . .	7
3.2.1 Time independent approaches . . . . .	7
3.2.2 Time dependent approaches . . . . .	9
3.2.3 Direct calculations of Reactions rate coefficients . . . . .	10
<b>4 Approximate Methods</b>	<b>12</b>
4.1 Reduced dimensionality quantum methods . . . . .	12
4.2 Mixed quantum and classical mechanics approaches . . . . .	13
4.3 The pure classical mechanics computational machinery . . . . .	13

---

\*Interesting discussions with the members of the Quantum Reactive Scattering (QRS) group of interest is acknowledged. The work illustrated here has been financially supported by MIUR, ASI, CNR and COST Chemistry (Action D23). Specific mention needs to be made to Italian Space Agency (Project ASI PQE2000) and the Italian MIUR FIRB Grid.it project (RBNE01KNFP) on High performance Grid Platforms and Tools, and to the MIUR CNR Strategic Project L 499/97-2000 on High performance Distributed Enabling Platforms.

## Report Documentation Page

*Form Approved  
OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE <b>01 JUL 2007</b>	2. REPORT TYPE <b>N/A</b>	3. DATES COVERED <b>-</b>			
4. TITLE AND SUBTITLE <b>Theory And ComputIng of Gas Phase Chemical Reactions : From Exact Quantum to Approximate Dynamical Treatments</b>					
5a. CONTRACT NUMBER					
5b. GRANT NUMBER					
5c. PROGRAM ELEMENT NUMBER					
6. AUTHOR(S)					
5d. PROJECT NUMBER					
5e. TASK NUMBER					
5f. WORK UNIT NUMBER					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Dipartimento di Chimica, Universit'a di Perugia, 06123 Perugia, Italy</b>					
8. PERFORMING ORGANIZATION REPORT NUMBER					
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)					
10. SPONSOR/MONITOR'S ACRONYM(S)					
11. SPONSOR/MONITOR'S REPORT NUMBER(S)					
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release, distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>See also ADM002058., The original document contains color images.</b>					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>UU</b>	18. NUMBER OF PAGES <b>24</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

<b>5 Advanced Computing</b>	<b>14</b>
5.1 Parallel computing . . . . .	15
5.2 Virtual organizations and Grid enabled applications . . . . .	16
5.3 The Grid implementation of a molecular science simulator . . . . .	17

## 1 Abstract

A realistic simulator of complex gas phase systems needs to handle accurately processes spanning different scales of space and time. The paper outlines the key theoretical and computational features of modern *a priori* treatments of the dynamics of elementary processes in these systems with particular emphasis on reactive processes.

## 2 Introduction

Realistic *a priori* simulations of rarefied gas flows are conceptually articulated into various blocks of operations characterized by different time and space scales (1). These blocks are concerned with the calculation of electronic structures, molecular collisions, collective (fluid-, electro-, etc.) dynamics and the averaging over randomly sampled variables to work out *a priori* estimates of observed and measurable properties of real situations (like the formation of shock waves in rarefied gas flows). As a matter of fact, these are the pillars on which the COST Chemistry Metalaboratory SIMBEX has developed the homonymous simulator of molecular beam experiments (1; 2) and the COMPCHEM (3) virtual organization (VO)<sup>1</sup> has assembled the Grid Enabled Molecular Simulator (GEMS) proposed for European funding as a specific targeted research project within the Activity IST-2005-2.5.6 Research Network Testbed of the 6th Framework Program (4).

The mentioned blocks of computations are carried out using separate suites of codes thanks to the fact that time and space scales associated with them are profoundly different. This means that the electronic energy of the molecular systems and the related wavefunctions can be calculated to a high level of accuracy from first principles using *ab initio* techniques within the Born-Oppenheimer (BO) (5) approximation (provided that, when necessary, at certain times of the process or at certain arrangements of the nuclei in the corresponding stationary scheme, the coupling between nuclear and electronic degrees of freedom is regained). In the BO approximation the electronic motion is assumed to depend only parametrically on the nuclear coordinates and the potential energy (electronic energy plus nuclear repulsions) of the nuclear motion is calculated (at several fixed geometries of the nuclei) using well consolidated quantum chemistry suites of programs (see for example refs. (6; 7; 8; 9; 10)) which will not be discussed here. The *ab initio* calculation of the electronic energy can be performed either time by time at the actual geometry of the moving system (*on-the-fly*) or, once for ever, at the first step of the computational procedure. In the latter case, *ab initio* calculations are performed for a large matrix of nuclear geometries (including those of the initial and final states of the considered process) and calculated values are then best-fitted by optimizing the parameters of a suitable potential energy functional.

The integration of the motion equations of the nuclei on the adopted potential energy surface (PES) allows to estimate the efficiency parameters of the elementary chemical processes being considered. These detailed efficiency parameters (with their temperature, energy, spatial, angular momentum, etc. dependencies) are the key set of input data needed to carry out realistic simulations of the kinetic regime of real gaseous systems.

In this paper we shall discuss on:

<sup>1</sup>COMPCHEM is registered at grid-it.cnaf.infn.it.

1. the development of few atom rigorous dynamical methods (section 2);
2. the extension of dynamical methods to larger systems (section 3);
3. the exploitation of the potentialities of advanced computing for Molecular Sciences calculations (section 4).

### 3 The *a priori* dynamical approach

From a theoretical point of view elementary chemical processes are many body problems concerning the encounter of two or more aggregates of electrons and nuclei. This allows to treat the molecules as deformable objects which can collide and break into parts and/or recombine. Simultaneous encounters of three or more objects are quite unlikely to occur (especially in the low pressure gas phase processes mentioned before). Accordingly, most of the theoretical and computational studies have focused on two body encounters (bi-molecular). After all, the treatment of multi-molecular encounters are usually rationalized by chaining relevant bi-molecular collisions (though, recently, they are increasingly dealt directly using classical molecular dynamics). For uni-molecular processes too the theoretical approach can be easily reconducted to that of half a bi-molecular encounter.

Typical efficiency parameters of elementary bi-molecular processes are either (thermal, state-specific and state-to-state) rate coefficients or integral (state-specific and state-to-state) cross sections. The thermal rate coefficient  $k(T)$  is formulated as (11)

$$k(T) = \sum_i \sum_f \frac{w_i \exp[-\epsilon_i/k_B T]}{Q_{int}(T)} k_{i,f}(T) \quad (1)$$

where  $i$  and  $f$  are respectively the initial and final internal states,  $w_i$  is the multiplicity of the initial internal state  $i$  including the nuclear spin symmetry,  $Q_{int}(T) = \sum_i w_i \exp[-\epsilon_i/k_B T]$  is the partition function associated with the initial internal states,  $k_B$  is the Boltzmann constant,  $\epsilon_i$  is the energy of state  $i$ ,  $k_{i,f}(T)$  is the state-to-state rate coefficient and  $T$  is the temperature of the system. The evaluation of the state-to-state rate coefficient can be reconducted to the calculation of the state-to-state cross section  $\sigma_{i,f}(E_{tr})$  using the following equation

$$k_{i,f}(T) = \left( \frac{8}{\pi \mu k_B^3 T^3} \right)^{1/2} \int_0^\infty E_{tr} \sigma_{i,f}(E_{tr}) e^{-E_{tr}/k_B T} dE_{tr} \quad (2)$$

if the energy distribution is of the Boltzmann type with  $\mu$  being the reduced mass of the system in its reactant arrangement and  $E_{tr}$  the translational energy. The state-to-state cross section ( $\sigma_{i,f}(E_{tr})$ ) can be calculated from the state-to-state cumulative reaction probability ( $P_{i,f}(E_{tr})$ ) as follows

$$\sigma_{i,f}(E_{tr}) = \frac{\pi}{k_i^2} P_{i,f}(E_{tr}) = \frac{\pi}{k_i^2} \sum_{J=0}^{\infty} (2J+1) P_{i,f}^J(E_{tr}) = \frac{\pi}{k_i^2} \sum_{J=0}^{\infty} (2J+1) \sum_{\Lambda=-J}^J P_{i,f}^{J\Lambda}(E_{tr}) \quad (3)$$

where  $k_i$  is the system wavenumber and the individual terms  $P_{i,f}^{J\Lambda}(E_{tr})$  of the right hand side member of expression 3 can be derived from the square modulus of the  $\mathbf{S}$  matrix

elements calculated at a given pair of total angular momentum quantum number  $J$  and helicity quantum number  $\Lambda$ .

### 3.1 Electronically adiabatic potential energy surfaces

As already mentioned, *a priori* calculations of state-specific and state-to-state probabilities of elementary chemical processes start from the consideration that the fixed nuclei electronic wavefunctions  $\Phi^e(\{\mathbf{w}\}; \{\mathbf{W}\})$  are a suitable basis set for the expansion of the system wavefunction  $Z(\{\mathbf{w}\}, \{\mathbf{W}\}, t)$  (in our notation  $\{\mathbf{w}\}$  and  $\{\mathbf{W}\}$  are the sets of electronic and nuclear position vectors, respectively, and  $t$  is time). The fixed nuclei electronic wavefunctions  $\Phi^e(\{\mathbf{w}\}; \{\mathbf{W}\})$  are the eigensolutions of the fixed nuclei electronic problem

$$\hat{H}_e(\{\mathbf{w}\}; \{\mathbf{W}\}) \Phi^e(\{\mathbf{w}\}; \{\mathbf{W}\}) = \mathcal{E}(\{\mathbf{W}\}) \Phi^e(\{\mathbf{w}\}; \{\mathbf{W}\}) \quad (4)$$

where  $\hat{H}_e$  is the time independent electronic Hamiltonian. The solution of eq. 4 provides us with the desired set of  $I$ th adiabatic electronic eigenvalues  $\mathcal{E}_I(\{\mathbf{W}\})$  for the molecular geometry and electronic wavefunction associated with it. The ensemble of the  $\mathcal{E}_I(\{\mathbf{W}\})$  values, once summed to the corresponding nuclear repulsion, represent a pointwise description of the PES ( $V_I(\{\mathbf{W}\})$ ) on which the motion of the  $N_{nucl}$  nuclei of the system takes place.

This gives us a means for solving the general time dependent Schrödinger equation of the system (12; 13)

$$i\hbar \frac{\partial}{\partial t} Z(\{\mathbf{w}\}, \{\mathbf{W}\}, t) = \hat{H}(\{\mathbf{w}\}, \{\mathbf{W}\}) Z(\{\mathbf{w}\}, \{\mathbf{W}\}, t) \quad (5)$$

(in eq. 5  $\hat{H} = \hat{H}_e + \hat{H}_N$  is the total many body Hamiltonian of the system whith  $\hat{H}_N$  being its nuclear component) by expanding  $Z$ , as already mentioned, in terms of the electronic eigenfunctions which parametrically depend on the position vectors of the nuclei as follows

$$Z(\{\mathbf{w}\}, \{\mathbf{W}\}, t) = \sum_I \Psi_I(\{\mathbf{W}\}, t) \Phi_I^e(\{\mathbf{w}\}; \{\mathbf{W}\}). \quad (6)$$

After averaging over the electronic coordinates, the resulting differential equations for the  $\Psi_I(\{\mathbf{W}\}, t)$  coefficients of the expansion contain some terms coupling nuclear and electronic degrees of freedom. As already mentioned, the BO decoupling scheme is usually applied at this point by assuming these terms to be negligible. As a result, for each electronic state  $I$  (hereafter, the index  $I$  will be dropped from the formalism because we shall confine our attention to the single electronically adiabatic PES BO regime) the calculations reduce to the following electronically adiabatic time dependent Schrödinger equation (12; 13; 14)

$$i\hbar \frac{\partial}{\partial t} \Psi(\{\mathbf{W}\}, t) = \hat{H}_N(\{\mathbf{W}\}) \Psi(\{\mathbf{W}\}, t) = \left[ \hat{T}_N(\{\mathbf{W}\}) + V(\{\mathbf{W}\}) \right] \Psi(\{\mathbf{W}\}, t) \quad (7)$$

where  $\hat{T}_N(\{\mathbf{W}\})$  is the nuclear kinetic operator.

In regions where the BO approximation breaks down, as is the case of closely spaced electronic eigenvalues, ad hoc treatments can be adopted which deal at the same time with different terms of the electronic functions manifolds (15). In these approaches the

elements coupling electronic and nuclear degrees of freedom are evaluated consistently using information coming from electronic structure calculations. Then resulting differential equations coupling different electronic states and wavefunctions are integrated using the same standard numerical algorithms available to integrate equations 7.

The crucial starting phase of the calculation of cross sections and rate coefficients is therefore the assemblage of a suitable PES. This task is most often highly demanding in terms of computing time and chemical ingenuity. For this reason the calculation of the potential energy values is increasingly tackled at an *on-the-fly ab initio* level. Yet, up to date, the most popular approaches are those based on a two step procedure. The first step consists of the collection of all the available (both *ab initio* and empirical) local or global data on the system interaction while the second step consists of their fitting using an appropriate functional form.

For small systems ( $N_{nucl} < 10$ ), as is the case of the majority of the elementary reactions considered up to date for theoretical dynamical studies (in some cases this is true also for larger systems if the complexity of the computational procedure is reduced by imposing suitable dynamical constraints), the above mentioned two step procedure is, indeed, the preferred one. The reason for this preference is the fact that the quality of present electronic structure calculations is often insufficient to guarantee an accurate reproduction of the interaction of the system over the full range of internuclear distances unless calculated *ab initio* values are adjusted using empirical considerations before undertaking the fitting. The functional representation to be used for fitting the PES of reactive systems is more difficult to formulate and most of the computations have been confined to systems made of three and four atoms (16; 17; 18; 19; 20).

The most popular functional forms used for this purpose are polynomials either in physical coordinates (17) (like the internuclear distances defined as  $r_{ij} = |\mathbf{W}_i - \mathbf{W}_j|$ ) or in bond order variables (21) (defined as exponentials of the displacement from equilibrium of the related internuclear distances). When using physical coordinates the polynomials need to be damped to avoid divergence at long range. Polynomial functionals are usually adopted within a Many Body Expansion approach to formulate the individual components of the expansion. Other simple functional forms are either derived from drastically simplified formulations of *ab initio* methods (22; 23) or from intuitive models (such as diatomic rotating potentials (24; 25)). Alternative approaches make use of local interpolation methods in which for each interval low order polynomials are employed and the value of related parameters are determined by imposing pointwise or switching continuity conditions. Similar approaches are also used for multi-surface treatments by fitting each surface using a functional form (except for methods directly providing multiple solutions).

When moving to complex systems it becomes more convenient to compose the PES by summing simple few body functions (stretches, bends, torsions, van der Waals, non-bonded interactions, etc) containing empirically determined parameters (force fields) (26). Parameters used by these approaches are transferable within the same family of systems. Usually these surfaces are scarcely suited to describe the making or breaking of bonds while they are better suited for conformational studies.

### 3.2 Quantum formalism for few body reactions

Quantum techniques based on the integration of eq. (7) are widely used to evaluate the observable properties of reactive elementary processes. This task is, nowadays, a largely routine work when dealing with atom diatom systems. From the integration of eq. (7) one can evaluate the elements of the scattering matrix  $\mathbf{S}$  whose square moduli are the elements of the probability matrix  $\mathbf{P}$  used in eq. 3.

For a few body isolated system (as is the case of laser free crossed molecular beam experiments) the motion of the center of mass of the system can be factored out to reduce the dimensionality of the problem (to six spatial coordinates in the atom diatom case that is the simplest prototype of elementary reactions) without introducing additional approximations.

Accordingly, for a three atom (A, B, C) system, the time dependent Schrödinger equation for the nuclei reads as

$$i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{R}_\tau, \mathbf{r}_\tau, t) = \left[ -\frac{\hbar^2}{2\mu} (\nabla_{\mathbf{R}_\tau}^2 + \nabla_{\mathbf{r}_\tau}^2) + V(R_\tau, r_\tau, \Theta_\tau) \right] \Psi(\mathbf{R}_\tau, \mathbf{r}_\tau, t) \quad (8)$$

where  $\mathbf{R}_\tau$  and  $\mathbf{r}_\tau$  are the mass scaled atom-diatom Jacobi vectors (of modulus  $R_\tau$  and  $r_\tau$ ),  $\Theta_\tau$  is the angle formed by  $\mathbf{R}_\tau$  and  $\mathbf{r}_\tau$  and  $\Psi(\mathbf{R}_\tau, \mathbf{r}_\tau, t)$  is the time dependent nuclear wavefunction. As usual, Jacobi coordinates are labeled after the arrangement  $\tau$  ( $\tau = 1, 2$  and 3 means A + BC, B + CA and C + BA respectively) to which they refer.

The dimensionality of the problem is further reduced if the Laplacian (the kinetic component of the Hamiltonian) is written in terms of angular momentum operators

$$-\frac{\hbar^2}{2\mu} (\nabla_{\mathbf{R}_\tau}^2 + \nabla_{\mathbf{r}_\tau}^2) = -\frac{\hbar^2}{2\mu} \left( \frac{1}{R_\tau} \frac{\partial^2}{\partial R_\tau^2} R_\tau + \frac{1}{r_\tau} \frac{\partial^2}{\partial r_\tau^2} r_\tau \right) + \frac{(\hat{J} - \hat{j}_\tau)^2}{2\mu R_\tau^2} + \frac{\hat{j}_\tau^2}{2\mu r_\tau^2} \quad (9)$$

with  $\hat{J}$  being the total angular momentum operator given by the sum of  $\hat{j}_\tau$  and  $\hat{l}_\tau$  (the rotational and the orbital angular momentum operators of the system, respectively). This makes it convenient to express the  $\Psi(\mathbf{R}_\tau, \mathbf{r}_\tau, t)$  wavefunction in terms of products of the  $\Psi^{JMp}(R_\tau, r_\tau, \Theta_\tau, t)$  partial waves (which are eigenfunctions of the eigenvalue  $J(J+1)$  of the total angular momentum operator  $\hat{J}^2$ , of its projection  $M$  on a space fixed (SF) reference axis and of the parity  $p$ ) in the internal coordinates  $R_\tau$ ,  $r_\tau$  and  $\Theta_\tau$ , and the appropriate spherical harmonics. For computational convenience to integrate the scattering equations from the reactant to the product arrangement one can also adopt a body fixed (BF) representation (in which the reference frame is allowed to rotate in order to have the  $z$  axis always aligned with the  $\mathbf{R}_\tau$  vector and the  $xz$  plane having a fixed orientation with respect to the molecular plane) where  $\Lambda$  is the projection of the total angular momentum  $\hat{J}$  on the BF  $z$  axis.

#### 3.2.1 Time independent approaches

To further reduce the dimensionality of the problem (this was the approach usually adopted in the past) the time variable can be factored out from the system wavefunction and a time independent formulation of the Schrödinger equation can be obtained without introducing additional approximations (this means that the system can be described by a stationary wave) (27). To integrate the stationary Schrödinger equation one

needs to define a particular coordinate (usually called reaction coordinate) by properly combining the internuclear distances or Jacobi coordinates to ensure a smooth switch from reactant oriented to product oriented arrangement coordinates.

The most popular reaction coordinate adopted in recent studies has been the hyper-radius  $\rho$  defined as  $\rho^2 = R_\tau^2 + r_\tau^2$  (this relationship holds for all arrangements because  $\rho$  is invariant under kinematic rotations). The hyperradius together with two hyperangles constitutes the set of hyperspherical coordinates. These coordinates can be defined in various ways depending on what arrangement is to be preferred. Here, for illustrative purposes, we make use of their APH (28) (democratic) version in which the hyperangles are  $\theta$  and  $\chi$  (the value of  $\chi$  depends on the choice of the reference geometry though for simplicity the related label is dropped here from the notation) and the partial wave equations take the form

$$\left[ \hat{T}_\rho + \hat{T}_h + \hat{T}_r + \hat{T}_c + V(\rho, \theta, \chi) \right] \Psi^{JMp}(\rho, \theta, \chi) = E \Psi^{JMp}(\rho, \theta, \chi) \quad (10)$$

where subscripts “ $h$ ”, “ $r$ ” and “ $c$ ” stand for “hypersphere”, “rotational,” and “Coriolis”, respectively and the operators  $\hat{T}_\rho$ ,  $\hat{T}_h$ ,  $\hat{T}_r$  and  $\hat{T}_c$  are formulated as:

$$\begin{aligned} \hat{T}_\rho &= -\frac{\hbar^2}{2\mu\rho^5} \frac{\partial}{\partial\rho} \rho^5 \frac{\partial}{\partial\rho}, \\ \hat{T}_h &= -\frac{\hbar^2}{2\mu\rho^2} \left( \frac{4}{\sin 2\theta} \frac{\partial}{\partial\theta} \sin 2\theta \frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\chi^2} \right), \\ \hat{T}_r &= A(\rho, \theta) J_x^2 + B(\rho, \theta) J_y^2 + C(\rho, \theta) J_z^2, \end{aligned}$$

and

$$\hat{T}_c = -\frac{i\hbar \cos\theta}{\mu\rho^2 \sin^2\theta} J_y \frac{\partial}{\partial\chi},$$

with  $A(\rho, \theta)$ ,  $B(\rho, \theta)$  and  $C(\rho, \theta)$  being defined as  $A^{-1}(\rho, \theta) = \mu\rho^2(1 + \sin\theta)$ ,  $B^{-1}(\rho, \theta) = 2\mu\rho^2 \sin^2\theta$ ,  $C^{-1}(\rho, \theta) = \mu\rho^2(1 - \sin\theta)$ .

Eq. (10) is integrated by segmenting the hyperradius in several sectors and expanding  $\Psi$  (within each sector  $i$ ) in terms of the  $\Phi_{t\Lambda}^{Jp}$  surface functions which are eigensolutions of the following equation

$$\left[ \hat{T}_h + \frac{15\hbar^2}{8\mu\rho_i^2} + (C - D) \hbar^2 \Lambda^2 + D\hbar^2 J(J + 1) + V(\rho_i, \theta, \chi) - \varepsilon_{t\Lambda}^{Jp}(\rho_i) \right] \Phi_{t\Lambda}^{Jp}(\theta, \chi; \rho_i) = 0 \quad (11)$$

with  $D = (A + B)/2$  (though the  $\Phi$  functions could be chosen also to be independent of  $J$  by setting  $J = 0$  in eq. 11 and regaining the  $J$  dependence during the integration over  $\rho$ ). Once the expansion is performed one gets the following set of equations to integrate over the hyperradius  $\rho$

$$\left[ \frac{\partial^2}{\partial\rho^2} + \frac{2\mu E}{\hbar^2} \right] \psi_{t\Lambda}^{Jpn}(\rho_i) = \frac{2\mu}{\hbar^2} \sum_{t'\Lambda'} \langle \Phi_{t\Lambda}^{Jp}(\theta, \chi; \rho_i) \hat{D}_{\Lambda M}^{Jp} | \hat{H}_i | \Phi_{t'\Lambda'}^{Jp}(\theta, \chi; \rho_i) \hat{D}_{\Lambda' M'}^{Jp} \rangle \psi_{t\Lambda}^{Jpn}(\rho_i) \quad (12)$$

where the internal Hamiltonian  $\hat{H}_i$  is defined as

$$\hat{H}_i = \hat{T}_h + \hat{T}_r + \hat{T}_c + \frac{15\hbar^2}{8\mu\rho^2} + V(\rho, \theta, \chi). \quad (13)$$

Accordingly, the computational procedure can be partitioned into three parts. Part one is devoted to the calculation of the  $\Phi$  functions over suitable fixed  $\rho$  grid points of  $\theta$  and  $\chi$  values. This part is dominated by the evaluation of (two dimensional for three atom, five dimensional for four atom systems) integral quadratures and eigenvalues finding of large real dense square matrices. The second part of the calculation consists in propagating the  $\psi_{t\Lambda}^{Jpn}(\rho_i)$  solutions. This part of the calculation is dominated by the inversion of large matrices having the same dimension as the number of channels. The third part is less computationally demanding and is devoted to the mapping of the asymptotic solution into the proper arrangement space and to the imposition of boundary conditions to work out the  $\mathbf{S}$  matrix.

Moving from three to four atom systems (say diatom-diatom) the number of internal coordinates doubles. A possible choice is the set of coordinates  $R$ ,  $r_1$ ,  $r_2$ ,  $\Theta_1$ ,  $\Theta_2$  and  $\phi$  with  $R$  being the distance between the centers of mass of the two diatoms,  $r_1$  and  $r_2$  being the two diatomic internuclear distances,  $\Theta_1$  and  $\Theta_2$  being the two planar angles formed by  $r_1$  and  $r_2$  with  $R$  and  $\phi$  being the dihedral angle formed by the  $(R, r_1)$  and  $(R, r_2)$  planes. Using these coordinates the Hamiltonian of the AB + CD system takes the form (29)

$$\hat{H}_N = -\frac{\hbar^2}{2\mu_4} \frac{\partial^2}{\partial R^2} + \hat{h}_1(r_1) + \hat{h}_2(r_2) + \frac{(\hat{J} - \hat{j}_{12})^2}{2\mu_4 R^2} + \frac{\hat{j}_1^2}{2\mu_1 r_1^2} + \frac{\hat{j}_2^2}{2\mu_2 r_2^2} + \Delta V \quad (14)$$

where  $\mu_4$  is the reduced mass of the AB and CD reduced masses ( $\mu_1$  and  $\mu_2$ , respectively),  $\hat{J}$  is, as already mentioned, the total angular momentum operator,  $\hat{j}_{12}$  is the sum of  $\hat{j}_1$  and  $\hat{j}_2$  which are the rotational angular momentum operators of AB and CD, respectively,  $\hat{h}_1(r_1)$  and  $\hat{h}_2(r_2)$  are the vibrational Hamiltonians of AB and CD, respectively, while  $\Delta V$  is the difference between the total interaction potential  $V(R, r_1, r_2, \Theta_1, \Theta_2, \phi)$  and the  $V(r_1)$  and  $V(r_2)$  vibrational potentials of  $\hat{h}_1(r_1)$  and  $\hat{h}_2(r_2)$ .

The increased complexity of the four atom Hamiltonian makes the definition of the reaction coordinate, the calculation of the sector basis functions and the switch from one arrangement to another (and therefore the solution of the time independent Schrödinger equation) very difficult. As a matter of fact, only recently significant advances have been made in describing the reaction coordinate of four atom systems by using row-orthonormal hyperspherical democratic coordinates made of a hyperradius and five hyperangles (30).

### 3.2.2 Time dependent approaches

The difficulty of handling the problem associated with the definition of a suitable smoothly evolving spatial continuity variable brings the discussion back to the decision of reducing the dimensionality of the calculation by factoring out the time dependence of the wavefunction. In this respect the simplicity obtained in formulating the computational machinery when using the Jacobi coordinate time dependent formalism is increasingly considered a suitable reward for keeping the extra variable of time in the formalism. As a matter of fact, in time dependent approaches one has the advantage of straightforwardly shaping the initial wavepacket. This is, in fact, chosen to correspond to a given reactant state or to a mixture of them, its component along the atom diatom coordinate  $R$  is formulated as a product of a gaussian wave ( $\exp[-\alpha(R - R_o)^2]$ ) times a phase shift factor ( $\exp[-ik_o(R - R_o)]$ ) associated with its inward traveling nature having an average

momentum  $k_o\hbar$ . Then the wavepacket can be mapped into any other set of arrangement coordinates of interest and let propagate in time by repeatedly applying the evolution operator  $\exp[-i\hat{H}\tau/\hbar]$ . Eventually, after the wavepacket has spread all over the whole accessible configuration space one can carry out its analysis at the product asymptotic line by expanding the cut of the wavefunction into the related diatomic wavefunctions (for atom-diatom systems). The time dependent coefficients of the expansion  $C_{v'j'\Lambda'}(t)$  read (31)

$$C_{v'j'\Lambda'}(t) = \int_r \int_{\Theta} \sin \Theta P_j^{\Lambda}(\Theta) \phi_{v'j'}(r) \Psi^{J\Lambda}(R = R_{\infty}, r, \Theta, t) d\Theta dr \quad (15)$$

when using the reactant Jacobi coordinates  $R$ ,  $r$  and  $\Theta$  (primed variables are for products, unprimed for reactants). In eq. (15)  $P_j^{\Lambda}(\Theta)$  is the normalized associated Legendre function of the angular part of the wavefunction. By performing a half Fourier transform of the  $C_{v'j'\Lambda'}(t)$  coefficients one gets the time independent (energy dependent)  $\mathbf{A}$  matrix whose elements read

$$A_{v'j'\Lambda'}(E) = \frac{1}{2\pi} \int_{t=0}^{\infty} \exp(iEt/\hbar) \cdot C_{v'j'\Lambda'}(t) dt. \quad (16)$$

From them one can easily determine the  $\mathbf{S}$  matrix elements whose square modulus (the probabilities) allow to calculate the atom-diatom cross section (31)

$$\sigma_{vj,v'j'}(E_{tr}) = \frac{\pi}{k_{vj}^2(2j+1)} \sum_J (2J+1) \sum_{p=\pm 1} \sum_{\Lambda,\Lambda'} |S_{vj\Lambda,v'j'\Lambda'}^J|^2. \quad (17)$$

### 3.2.3 Direct calculations of Reactions rate coefficients

The simplicity of the time dependent method has also facilitated its extension to larger systems (29; 32). However, when one is interested in evaluating the (less detailed) thermal rate coefficient there is no need to carry out in the computation all the details embodied in the  $\mathbf{S}$ -matrix. After all, the value of the rate coefficient is rather insensitive to the detailed structure of the whole PES while it is strongly dependent on the shape of the reactant side of the saddle and in particular on the height and the width of the reaction barrier (33).

This has motivated the formulation of the rate coefficient directly in terms of the cumulative reaction probability  $N(E)$  defined as

$$N(E) = \sum_i \sum_f P_{i,f}(E). \quad (18)$$

and that does not refer to any asymptotic state and depends only on the dynamics of the system in the vicinity of the reaction barrier. In terms of  $N(E)$  the rate coefficient is then formulated as

$$k(T) = \frac{1}{hQ_{tr}(T)Q_{int}(T)} \int_0^{\infty} N(E) e^{-E/k_B T} dE \quad (19)$$

where  $Q_{tr}(T)$  is equal to  $(2\pi\mu k_B T)^{3/2}/h^3$ . This means that the efficiency of the reactive process is expressed in terms of the fraction of the system wavefunction left over after

projecting out all its components which do not have outgoing character in the product asymptotic region. The projection operator can be given the form

$$P_{prod} = \lim_{t \rightarrow \infty} e^{i\hat{H}t} \theta(R - R_0) e^{-i\hat{H}t}. \quad (20)$$

Accordingly, the wavefunction can be propagated forward (out of an arbitrarily located dividing surface) infinitely in time. Then components not located on the product side of the dividing surface are projected out and the remaining wavefunction is propagated backward in time. In this approach, the rate constant  $k(T)$  can be formulated as:

$$k(T) = \frac{1}{Q_{tr}(T)Q_{int}(T)} \lim_{t \rightarrow \infty} \text{tr}(\hat{F} e^{-\hat{H}/k_B T} e^{i\hat{H}t} \hat{h} e^{-i\hat{H}t}) \quad (21)$$

when such a limit exists. In equation 21,  $\hat{h}$  can be chosen to be any operator discriminating between reactants and products and  $\hat{F}$  any operator measuring the flux from reactants to products (in Eq. (21) the following correlation function

$$C_{fp}(t) = \text{tr}(e^{-\frac{\hat{H}}{2kT}} \hat{F} e^{-\frac{\hat{H}}{2kT}} e^{i\hat{H}t} \hat{h} e^{-i\hat{H}t}) \quad (22)$$

of the flux-position type is used). As an alternative use can be made of the following correlation function

$$C_{ff}(t) = \text{tr}(e^{-\frac{\hat{H}}{2kT}} \hat{F} e^{-\frac{\hat{H}}{2kT}} e^{i\hat{H}t} \hat{F} e^{-i\hat{H}t}) \quad (23)$$

of the flux-flux type with  $C_{ff}(t)$  being the time derivative of  $C_{fp}(t)$  and  $\hat{F}$  the time derivative of  $\hat{h}$  in the Heisenberg picture. Accordingly, the thermal rate coefficient takes the form

$$k(T) = \frac{1}{Q_{tr}(T)Q_{int}(T)} \lim_{t \rightarrow \infty} C_{fp}(t) = \frac{1}{Q_{tr}(T)Q_{int}(T)} \int_0^{\infty} C_{ff}(t) dt \quad (24)$$

and the cumulative reaction probability becomes (34; 35):

$$\begin{aligned} N(E) &= 2\pi^2 \text{tr} \left( \hat{F} \delta(\hat{H} - E) \hat{F} \delta(\hat{H} - E) \right) \\ &= \frac{1}{2} e^{2E/k_B T} \sum_{f_T} \sum_{f'_T} f_T f'_T \left| \int e^{iEt} \langle f_T | e^{-i\hat{H}t} | f'_T \rangle dt \right|^2 \end{aligned} \quad (25)$$

where the evaluation of the trace (whose detailed calculation would imply the propagation of the whole set of basis functions) has been obtained in terms of the eigenstates  $f_T$  of the thermal flux operator  $\hat{F}_T$  defined as

$$\hat{F}_T = e^{-\hat{H}/2k_B T} \hat{F} e^{-\hat{H}/2k_B T} \quad (26)$$

which need only a small number of applications of  $\hat{F}_T$  on a sample wavefunction.

An interesting analogy between the classical and the quantum description of the rate coefficient can be obtained if one factorizes the correlation factor into a static and a dynamic component. The dynamic component provides, in fact, in the  $t \rightarrow \infty$  limit, a description of the amount of  $\hat{F}_T$  eigenstates ending up on the product side of the dividing surface. Different ways of exploiting these ideas to the end of carrying out the actual calculations of the value of the thermal rate coefficients and of cumulative reaction probabilities, are given in refs. (36; 37; 38; 39).

## 4 Approximate Methods

The practical impossibility of carrying out exact calculations for more complex systems has prompted the development of computational procedures based on various decoupling schemes effective in reducing the dimensionality of the problem.

### 4.1 Reduced dimensionality quantum methods

The goal of working out approximate formulations of the cross section and of the rate coefficient was first accomplished for atom diatom systems by introducing decoupling schemes of the kind energy sudden (40), centrifugal sudden (41) and infinite order sudden (42). In these decoupling schemes either diatomic rotations or atom diatom orbiting (or both) are treated in an approximate way. For systems made of four or more atoms the coupling of the various degrees of freedom, while significantly increasing the complexity of the dynamical treatment, most often plays a negligible role in determining the reaction outcome.

This has allowed a split of the dynamical treatment of strongly coupled degrees of freedom from weakly coupled ones. Strongly interacting degrees of freedom are treated rigorously while weakly interacting ones are treated approximately. The most popular approximations are based either on adiabatic assumptions or on the parametrization of some variables. As an example, in the rotating bond approximation (RBA) of diatom-diatom non linear collisions, two radial coordinates and one bending angle are explicitly treated while the other three degrees of freedom are kept frozen (43). The overall result is then worked out by averaging quantities calculated at the different values of the frozen variables. In the adiabatic bend approximation (ABA) (44; 45), instead, the three radial coordinates are treated explicitly while the three bending angles are treated adiabatically. This means that the overall wavefunction is factored out and eigenvalues associated with the effect of the terms of the Hamiltonian in the related coordinates are used to obtain the effective Hamiltonian for the coordinates to be treated exactly.

As already mentioned the calculation of dynamical properties can be performed by determining the cumulative probability which can be estimated in an approximate way using a transition state (TS) schematization of the reactive process. In this view the role played by the overall rotation of the system is that of shifting in energy (of a quantity  $E_{shift}$  related to the rotational energy of the system at the TS geometry) the reaction probability  $P_{i,f}^J(E_{tr})$ . This approximation, usually called  $J$ -shifting (45), links the value of the state-to-state probability calculated at a given value of the total angular momentum quantum number  $J$  to that calculated at a reference value ( $J_{ref}$ ). The relationship used is

$$P_{v,j,v'j'}^J(E_{tr}) = P_{v,j,v'j'}^{J_{ref}}(E_{tr} - E_{shift}) \quad (27)$$

where  $E_{shift}$  is the difference in energy between the overall rotational eigenstates  $J$  and  $J_{ref}$  at the transition state and  $J_{ref}$  is usually taken equal to zero. A more elaborated method linearly interpolates  $P_{i,f}^J(E_{tr})$  between the probabilities calculated at the two  $J_{ref}$  values most closely sandwiching the actual  $J$  value of the calculation.

## 4.2 Mixed quantum and classical mechanics approaches

Approaches alternative to quantum calculations are those based on classical mechanics. Several of these approaches use classical mechanics formulations to describe slower motions and quantum mechanics formulations to describe the remaining (faster) degrees of freedom. Related schemes have been applied to divide electronic from nuclear motion and to formulate widely general scaling procedures (47; 48). Some of these mixed quantum-classical approximations have been derived in a rigorous way in refs. (49; 50) by introducing a specific basis-set of the orthogonal polynomial type centered at a "classical" trajectory and expanded around it as the dynamics evolves in time. In this case the forces appearing in the classical equations of motion are not the usual Newtonian ones but more general forces usually called "quantum forces".

The chosen basis can be used in combination with the usual time independent basis functions or collocated representations. This allows to work out mixed quantum classical approaches and to monitor the quantum classical correlation and measure the accuracy of classical path treatments.

When systems become very large, the computational procedures need to be simplified further also at the level of the calculation of the electronic energies (in this case, in fact, the accuracy of the calculated electronic wave function becomes intrinsically poor (51)). As a result, the choice of election is to combine the use of classical mechanics for treating the dynamics of the nuclei with the adoption of approximate schemes to calculate electronic energy (like in density functional approaches (51)) and greatly simplify the *on-the-fly* (52) computational machinery.

These *Ab-initio* Molecular Dynamics simulations are presently (53; 54) applied to the study of physico-chemical properties, such as reactivity and dynamic relaxation, of several systems (55). Hybrid procedures linking Quantum-Mechanical parameterized descriptions of the 'active sites' of the molecule with a Molecular Mechanics description of its inactive framework are highly popular. These Quantum Mechanics Molecular Mechanics (QMMM) treatments (56) show advantages and limitations. They still lead, in fact, to dynamical results which cannot be inferred by a mere static analysis of the features of the potential energy surface though they do not treat most of them accurately.

At various stages of a trajectory calculation one has the possibility of treating semiclassically those degrees of freedom for which a classical approach is inadequate by associating to them a semiclassical wave depending on the classical action accumulated along the classical path. This allows to regain concepts like flux, interference, resonance and tunneling within a trajectory framework and reproduce quite closely some quantum features of the results.

## 4.3 The pure classical mechanics computational machinery

Purely classical mechanics approaches found the dynamical treatment on the integration of classical trajectories (CT). For this reason they are called quasiclassical or QCT when initial and final states are in some way discretized. CT methods assume that the nuclei involved in a chemical reaction obey classical mechanics and roll as point mass particles on the potential energy surface of the system. Accordingly,  $H_N^{CT}(\mathbf{R})$ , the classical analogue

of the quantum nuclear Hamiltonian  $H_N(\mathbf{R})$  of eq. (7), reads

$$H_N^{CT}(\{\mathbf{R}\}) = \sum_k^{N_{nucl}} \frac{p_{k,x_{R_k}}^2 + p_{k,y_{R_k}}^2 + p_{k,z_{R_k}}^2}{2m_k} + V(\{\mathbf{R}\}) \quad (28)$$

and the evolution of the system is followed by integrating the equations of classical mechanics starting from different sets of initial conditions (for atom diatom systems initial conditions are given by the vibrational number  $\nu$ , the rotational number  $j$ , the translational energy  $E_{tr}$ , the impact parameter  $b$ , the elongation of the diatom, the phase of the rotating diatom, the angles formed by the rotational angular momentum and the velocity with the molecular plane). Various formulations of the classical particles motion equations can be given. In the widely used Hamilton's version they read as

$$\begin{aligned} \frac{ds_{R_k}}{dt} &= \frac{\partial H_N}{\partial p_{s_{R_k}}} \\ \frac{dp_{s_{R_k}}}{dt} &= -\frac{\partial H_N}{\partial s_{R_k}} \end{aligned}$$

for each cartesian projection  $s_{R_k}$  (of the position vector  $\mathbf{R}_k$ ) and  $p_{s_{R_k}}$  (of the momentum vector  $\mathbf{p}_k$ ) of each atom  $k$  of mass  $m_k$  of the molecular system. The equations are integrated starting from one set of allowed initial conditions of the reactants in state  $i$  and are stopped either when the maximum number of interactions steps has been reached or when an asymptotic geometry of the products has been reached. As already mentioned, discrete features of quantum results are then enforced in QCT ones by arbitrarily discretizing the energy of classical bound motions. Usually for atom diatom systems this means that the counter  $N_{\nu j \nu' j'}$  (associated with trajectories starting from  $\nu j$  and ending with a classical vibrotational energy closer to that of the  $\nu' j'$  state than to any other one) is incremented by a unit. When all planned  $N_{tj}$  trajectories are integrated  $P_{\nu j, \nu' j'}$  is set equal to  $N_{\nu j, \nu' j'}/N_{tj}$  and  $\sigma_{\nu j, \nu' j'}$  is set equal to  $\pi b_{max} P_{\nu j, \nu' j'}$  (with  $b_{max}$  being the maximum value used for the impact parameter). CT and QCT approaches can often provide estimates of rate coefficients, cross sections, angular distributions and reaction probabilities of reasonable accuracy. Moreover, they allow a pictorial view of the mechanisms governing chemical reactions. Obviously, the CT method is an approximation to the nuclear motion and it becomes more accurate when quantum effects are negligible (as is the case of heavy nuclei, large collision energies and highly averaged reactive properties).

## 5 Advanced Computing

As it has been already mentioned, the demand for computer resources prompted by chemical reactivity calculations and related realistic a priori simulations is as high as that of other grand challenges of modern computational science. To guarantee the computational feasibility of these applications one has to resort to the exploitation of the innovative features of parallel and distributed computers by decomposing the problem in simpler decoupled subproblems and distributing the resulting independent blocks of the codes for execution on a large quantity of processing elements. The application of some decompositions based on physical considerations (like separating the electronic structure

calculation, factorizing the time variable, disentangling the center of mass motion and performing a partial wave expansion) have been already discussed before. However, when dealing with large systems and complex applications one has to effectively exploit the innovative features of parallel and distributed computing by carrying out a decomposition of the problem at an algorithmic level.

## 5.1 Parallel computing

To effectively tackle the problem of parallel restructuring a computational application the following aspects need to be specifically considered (57; 58; 59)

1. the key algorithms need to be optimized for parallel execution,
2. existing software modules need to be integrated efficiently,
3. different programming models and languages need to be used at various levels,
4. retrieving and storing of data structures need to be reorganized and streamed,
5. performances of the adopted articulation need to be measured under different conditions for its improvement.

This particularly difficult and time consuming job is usually carried out by reorganizing the relevant suites of codes and by inserting the appropriate directives and commands chosen among those of the most popular parallelization libraries (60). However, the need for ensuring both reliability and standardization on one side and the difficulty of keeping the pace of the continuous evolution of architectures and simulation techniques on the other side have made it necessary to produce tools guaranteeing the automatic or semi-automatic portability of applications (also in the sense of performance portability) across computing platforms.

Significant progress along this direction has been made using structured environments. A typical structured environment useful for a semi-automatic parallelization of the applications is SKIE (61). SKIE is an integrated environment providing a new application oriented set of instruments allowing the rapid development and prototyping of applications. Such an environment is based on some optimized and ready-to-use parallel structures, called *skeletons*. The skeletons can embody sections of codes allowing so far an extended reuse of the existing sequential (written in the most popular high level sequential languages) or parallel programs by encapsulating them in modules. Examples of skeletons are *processor farms* (a pool of worker processes computes a pool of independent tasks); *pipelines*, (different processes carry out in a sequence the various phases of the computation); *map data-parallel computations* (all the elements of a data structure are updated or computed at the same time). These structures are handled using a coordination language (CL) called SKIE-CL and can be utilized to coordinate and connect any sequential or parallel module encapsulated using a SKIE-CL wrapper. The wrapper ensures that parameter passing and data representation are consistent among the modules composing a parallel application. SKIE-CL makes use of instruments like *control*, *stream parallel* and *data parallel* though it accepts also in the input and output parameter lists all the usual basic types of variables (integers, real, etc), records, and multi-dimensional arrays.

In addition, SKIE-CL provides a *stream* data type which allows programmers to generate or consume sequences of items of indefinite length. The peculiar feature of these patterns is the fact that they can be freely composed "a la carte" to build complex structures.

SKIE automatically generates also an optimized implementation of the skeleton composition. This means that when using a SKIE skeleton the support not only generates automatically the code needed for parallel interaction, but it also optimizes the resources allocated to each skeleton, decides the best granularity of computation and locates inefficiencies in the global structure.

Going to a more abstract level skeletons can be generalized practically to any form (or combination of forms) and size. This is indeed the key feature of the coordination language ASSIST (62; 63) that is an evolution of SKIE. ASSIST is made of a graph whose nodes are parallel (*parmod*) and sequential (*seqmod*) modules. The arches of the graph represent streams of data. Interaction among parmods can occur also via shared objects (data, memory regions, functions, etc.) for which ASSIST provides a primitive access mechanism by implementing or emulating a shared memory access. Non primitive accesses are instead provided for external objects (like DSM libraries, CORBA servers, etc.) for which the user has to take care of the access, synchronization and consistency via directives to be inserted in the code. Activities in different parmods can be parallel or concurrent (like, for example, in a pipeline). Parallel or concurrent activities can take place also within the same parmod. They can be either farm-like or data-parallel-like (or mixed in a non-deterministic way depending on the structure of the data and the status of the computation). This means that ASSIST allows the following two hierarchical levels: among various parmods and within the same parmod. The first level describes a graph of the data flow type while the second can describe computations both of a data parallel domain decomposition type and of a functional replication farm type. ASSIST is also scarcely invasive since the computation is described using a set of procedures wrapping the sequential user code and organized in the parallel fashion specified by the used parmod type.

ASSIST has been already used to parallelize time dependent reactive scattering applications. Applications implemented using ASSIST and its libraries (see ref. (64)) have shown to clearly outperform their versions implemented using MPI. Similar studies have been carried out for time independent applications separately for the surface functions calculations and the propagation along the reaction coordinates. Tests performed using MPI on an eigenvalue finding routine are illustrated in ref. (65)

## 5.2 Virtual organizations and Grid enabled applications

Moving to realistic simulations of complex chemical systems, however, parallelism on a single (no matter how powerful) machine is still insufficient. As a matter of fact, the emerging computing paradigm is the computing Grid (66; 67). The Grid offers the possibility of solving complex problems using a (preferably very large) set of distributed computers as a single unified computing resource. To this end the Grids enable the sharing, selection, and aggregation of a wide variety of geographically dispersed resources ranging from PCs to supercomputers, storage systems, data sources and specialized devices. These may be owned by different organizations and work for completely different purposes within the virtual organization. Accordingly a Grid can be viewed as a seamless

integrated computational platform.

Moreover, the Grid is by definition at the same time both a collaborative environment and a market in which the users interact to find a solution for their problems and/or to offer their solutions to other users. Just to mention a few services which can be "traded" on the Grid we can list: Computational services, Data services, Application services, Information services and Knowledge services. This makes Grid applications intrinsically large scale and multidisciplinary. Obviously, in order to make the Grid computing effective, a large number of tools concerning security, resource allocation, costs management, information flux, software development, process execution, resource aggregation and scheduling need to be implemented. As a matter of fact the Grid is able to use in a synergistic way resources of all kinds including those which cannot be physically replicated in a single site. This means that one can easily scale up computing cycles as well as competences regardless of their location to work in a coordinated fashion. In summary, the Grid is a virtual place for composition of multiple administrative domains and autonomies to work concertedly on a large variety of heterogeneous machines and problems which naturally scales up from a few integrated resources to very many of them a smooth, dynamic, adaptable and interoperable way.

The European Grid of election is the already mentioned EGEE (68). Within EGEE a memorandum of understanding has been signed last March with the COMPCHEM virtual organization to foster the development of molecular science complex simulations. The mapping of a complex application on the computing grid is, indeed, a difficult task. Such a process is not simple because there is not a unique correspondence between the variation of physical and mathematical parameters of a complex computational application and the distribution on the grid of the computing blocks. This is not only due to the fact that a variation of the computational parameters alters the relative importance of the various computational blocks and of the related demand of computational resources but also to the fact that the support of the grid infrastructure at run time is not deterministic (69). For this reason it is vital to work out data graphs and build workflow managers allowing a proper independent handling of the computing blocks at various levels of distribution. The first step of this process is, therefore, the breaking of the computational procedure into independent or loosely coupled computing blocks. In fact, the singling out of independent (or almost independent) computational tasks is propedeutic to the design of any concurrent organization of the relevant computer programs.

### **5.3 The Grid implementation of a molecular science simulator**

The particular Grid enabled application considered as a study case by COMPCHEM is, as it has been already mentioned, GEMS the Grid Enabled Molecular Simulator. A demo version of GEMS (GEMS.0) has been already implemented on the production Grid of EGEE and presented at the first EGEE review workshop (70). GEMS.0 is derived from the SIMBEX simulator (see ref. (1)) developed by the homonymous working group of the COST Chemistry Action D23 (Metalaboratories for complex computational applications in Chemistry) (71). It takes care of evaluating the cross sections and the product distributions (plus some non observable quantities) of a crossed molecular beam atom-diatom experiment using a quasiclassical approach. In a quasiclassical approach, the observable properties of a scattering experiment are determined by performing a multidimensional

integral over the unselected variables of the experiment. The kernel of the integral is given by a boolean function whose value depends on the result of the integration of the Hamilton equations of the molecular system.

This means that one has to integrate large batches (say  $N_{tj}$ ) of classical trajectories for which some of the initial conditions are randomly selected. The calculation of each individual trajectory implies the integration of the equations of motion of the  $N_{nucl}$  atoms composing the system. This means that the simulation has to generate, possibly in a deterministic way, a large number of subsets of pseudorandom numbers (with each subset determining the initial conditions of a given trajectory). The overall workflow of GEMS.0 is, therefore, articulated into:

1. a first part defining the parameters of the calculation, computing quantities of general use and assembling the information needed for the calculation of the potential,
2. a final part performing the averaging of the calculated quantities and carrying out the graphical elaboration of the properties to be rendered possibly in real time on the virtual monitors,
3. a central (the key) part iterating over the distribution to the workers of the trajectories to be integrated and to the recollection of the related results to update the reactive probability.

The central part of GEMS.0 is the kernel of the calculation that can be efficiently distributed on the Grid using a task-farm scheme. The distribution can take place by assigning the integration of individual (or blocks of) trajectories to the worker nodes after generating  $initial_{seed}$  (the first seed of the subset needed by each trajectory or subset of trajectories) at Master process level. In GEMS.0, in order to keep the generation of  $initial_{seed}$  as much deterministic as possible, it has been chosen to perform the iteration on the individual trajectories. Accordingly, to integrate  $N_{tj}$  trajectories on a Grid made of  $N_{node}$  nodes ( $N_{node}$  is assumed to be smaller than  $N_{tj}$ ) the Master process (see Fig. 1) generates and SENDS out, for the first  $N_{node}$  iterations, the initial seed of each trajectory without waiting for the result of the integration to be returned. For the subsequent iterations the  $initial_{seed}$  of each trajectory is sent out only after one of the worker nodes ( $any_{node}$ ) has sent back its result. After the sending to the worker nodes of  $initial_{seed}$  for the  $N_{tj}$  trajectories is completed the Master process still needs to iterate over the  $N_{node}$  nodes to collect the results of the remaining trajectories and send a conventional signal (we chose to adopt for this purpose a negative value of the initial seed) to stop the work at worker process level.

As sketched in Fig. 2 the worker process devoted to the integration of each individual trajectory receives at the beginning (once for ever) the general information. Then it iterates on receiving  $initial_{seed}$  of the subset necessary to generate the trajectory initial conditions. After receiving the trajectory initial seed the worker process, if the seed is not negative, generates the remaining pseudorandom numbers of the needed subset to work out either at random (fully or partially according to the chosen distribution) or the specific initial value of the parameters of the trajectory like  $E_{tr}$ ,  $J$ , the velocity orientation, the diatomic internuclear distance, orientation and position as well as its vibrotational ( $v_j$ ) quantum state. The iteration on time then starts to integrate the trajectory from

```

Read input data:  $v, j, E_{tr}, \Delta t_o$ , error, maxstep,  $N_{tj} \dots$ 
Perform preliminary calculations
BROADCAST to the worker nodes the needed general information
Generate  $initial_{seed}$  for the first trajectory
 $i_{node}=0$ 
Do  $i_{tj} = 1, N_{tj}$ 
   $i_{node}=i_{node}+1$ 
  Generate next  $initial_{seed}$  of the of the pseudorandom series
  IF  $i_{node} > N_{node}$  THEN
    RECEIVE from  $any_{node}$  the result and update  $P_{vj,v'j'}$ 
    SEND the  $initial_{seed}$  to  $any_{node}$ 
    ELSE SEND the  $initial_{seed}$  to  $i_{node}$ 
  ENDIF
EndDo  $i_{tj}$ 
Do  $i_{node} = 1, N_{node}$ 
  RECEIVE from  $any_{node}$  the result and updated  $P_{vj,v'j'}$ 
  SEND to  $any_{node}$  a negative seed to stop activities
EndDo  $i_{node}$ 

```

Figure 1: Pseudocode of the trajectory Master program.

an atom-diatom distance large enough to consider the system in its asymptotic reactant arrangement. The integration is eventually terminated either when one of the atom-atom distances has reached a value large enough to consider again the arrangement as asymptotic or when the maximum number of integration steps (*maxstep*) has been reached. If at the integration end point the system has reached the product asymptote the  $P_{vj,v'j'}^J$  probability of the product diatom vibrational ( $v'j'$ ) quantum state closest in energy to the computed classical value is updated.

In GEMS.0 the PES is assumed to be of the LEPS type and the value of its parameters to be available from a library (they may have been already calculated using two other computational procedures, SUPSIM and FITTING, implemented in our Laboratory (72; 73)). The integration of a trajectory can be disposed to any computing node of the grid while the integration outcome is accumulated by the master process sketched in Fig. 1 by updating the value of the related quasiclassical probability  $P_{vj,v'j'}$  which can be displayed to the user on a virtual monitor.

Test runs of GEMS.0 (70) performed as EGEE (68) demonstrations have shown that the simulator is highly suitable for a distribution on the heterogenous networked environment of the Grid. In these runs from several thousands to millions or even billions of trajectories (to this end particular attention has to be paid to the generation of the pseudorandom sequence) can be run depending on how much the considered event is likely to occur. The main feature of GEMS.0 is that of being cpu bound (this is in general true even for systems larger than atom diatom ones). The implementation of GEMS.0 has impacted EGEE in two ways. The first of them is related to the specific requests of GEMS.0 in terms of infrastructure, middleware and services. The second is related to the characteristics that molecular simulations need to possess in order to be suitable for Grid

```

RECEIVE from the master node the needed general information
[*] RECEIVE the trajectory initialseed
IF seed < 0 STOP
Generate the needed subset of pseudorandom numbers
Calculate initial values of the integration variables
t = 0
tstep =  $\Delta t_o$ 
Do it = 1, maxstep
    t = t + tstep
    Perform the time step integration
    IF Energy and total angular momentum are conserved THEN
        IF an asymptotic arrangement has been reached THEN
            perform the asymptotic analysis
            SEND results to Master
            GOTO [*]
        ENDIF
    ELSE t = t - tstep and reduce tstep
    ENDIF
ENDDO it
GOTO [*]

```

Figure 2: Pseudocode of the trajectory worker program.

implementation.

As to the first item it has become apparent that molecular simulations force the Grid to better exploit various levels of parallelization and distribution using an appropriate workflow computational procedures of different nature and origin. In other words, GEMS requests EGEE to focus more on an effective coordination of brainware and knowledge.

As to the second item it has also become apparent that EGEE prefers direct calculations (with respect to data transfers) and coarse grain granularity schemes.

## References

- [1] Gervasi, O., Laganà, A., 2004, Future Generation Computer Systems, 20(5), 703.
- [2] O. Gervasi, D. Cicoria, A. Laganà and R. Baraglia, Pixel, 10, 19 (1944).
- [3] NA4: Memorandum of understanding for the EGEE applications, March 2005.
- [4] Laganà, A., GEMS-testbed: Grid Enabled Molecular Simulator, an EGEE testbed devoted to on demand interoperable cooperative simulations for Molecular Science Applications, proposal n. IST035244 GEMS-testbed of the 6th Framework Program.

- [5] M. Born and J. R. Oppenheimer, Ann. Phys. **84**, 457 (1927).
- [6] <http://www.molpro.net/>
- [7] <http://www.msg.ameslab.gov/GAMESS/GAMESS.html>
- [8] <http://www.gaussian.conv/g03.htm>
- [9] <http://www.emsl.pnl.gov/docs/nwchem>
- [10] [http://www.itc.univie.ac.at/ hans/Columbus/columbus\\_main.html](http://www.itc.univie.ac.at/ hans/Columbus/columbus_main.html)
- [11] P.L. Houston, Chemical kinetics and reaction dynamics, McGraw Hill, Boston 2001, ISBN 007-243537-2
- [12] M. Baer, The theory of chemical reaction dynamics CRC Press, Boca Raton, FL, 1985
- [13] D.C. Clary, The theory of chemical reaction dynamics, D. Reidel, Boston, 1986.
- [14] W.H. Lester Jr, The N coupled channel problem in Dynamics of Molecular Collisions, Plenum Press, New York, 1976.
- [15] J.C. Tully, Non Adiabatic Processes in Molecular Collisions in Dynamics of Molecular Collisions, Plenum Press, New York, 1976.
- [16] G. C. Schatz, Rev. Mod. Phys. **61**, 669 (1989); G. C. Schatz, Lect. Not. Chem. **75**, 15 (2000).
- [17] J. N. Murrell, S. Carter, S. C. Farantos, P. Huxley, and A. J. C. Varandas, *Molecular Potential Energy Functions* (Wiley, Chichester, 1984).
- [18] A. J. C. Varandas, Adv. Chem. Phys. **74**, 255 (1988).
- [19] A. J. C. Varandas, J. Molec. Str. (Theochem) **166**, 59 (1988).
- [20] D. G. Truhlar, R. Steckler, and M. S. Gordon, Chem. Rev. **87**, 217 (1987).
- [21] E. Garcia and A. Laganà, Mol. Phys **56**, 629 (1985).
- [22] F. O. Ellison, J. Am. Chem. Soc. **85**, 3540 (1963).
- [23] P. J. Kuntz, Ber. Bunsenges. Phys. Chem. **86**, 367 (1982).
- [24] A. Laganà, G. Ochoa de Aspuru, and E. Garcia, J. Chem. Phys. **108**, 3886 (1998).
- [25] G. Ochoa de Aspuru and D. C. Clary, J. Phys. Chem. A **102**, 9631 (1998).
- [26] U. Burkert and N.L. Allinger, *Molecular Mechanics* (ACS Monograph Series, Washington DC, 1982).

- [27] R. G. Newton *Scattering theory of waves and particles* (Springer-Verlag, New York, 1982).
- [28] R. T Pack and G. A. Parker, *J. Chem. Phys.* **87**, 3888 (1987).
- [29] D.H. Zhang, M. Yang, S. Lee, M. Collins, First principle quantum dynamical study of four atom reactions in Modern trends in chemical reaction dynamics Experiment and theory (Part II) *Adv. Series in Phys. Chem.*, X. Yang and K. Liu Ed., World Scientific, Singapore 2004.
- [30] A. Kuppermann, *J. Phys. Chem.* **100**, 2621 (1996); *J. Phys. Chem.* 1997, 101, 6368; *J. Phys. Chem.* **108**, 8894 (2004)
- [31] G. G. Balint-Kurti, in *International review of Science*, Series II, Vol. 1. Eds. A.D. Buckingham and C.A. Coulson (Butterworths), page 286 (1975).
- [32] D.M. Medvedev, E.M. Goldfield, S.K. Gray, *Comp. Phys. Comm* 166, 94 (2005).
- [33] U. Manthe, *Lecture Note in Chemistry* **75**, 130 (2000).
- [34] W. H. Miller, S. D. Schwartz, and J. W. Tromp, *J. Chem. Phys.* **79**, 4889 (1983).
- [35] F. Matzkies and U. Manthe, *J. Chem. Phys.* **110**, 88 (1999).
- [36] D. H. Zhang and J. C. Light, *J. Chem. Phys.* **104**, 6184 (1996).
- [37] S. M. Miller and T. Carrington, *Chem. Phys. Lett.* **267**, 417 (1997).
- [38] W. H. Thompson and W. H. Miller, *J. Chem. Phys.* **106**, 142 (1997).
- [39] F. Matzkies and U. Manthe, *J. Chem. Phys.* **108**, 4828 (1998).
- [40] D.S. Chase, *Phys. Rev.* 104, 838 (1956).
- [41] R.T Pack *J. Chem. Phys.* 60, 633 (1974); P. McGuire and D.J. Kouri, *JCP* 60, 248 (1974).
- [42] M. Baer, E. Garcia, A. Laganà, O. Gervasi, *Chem. Phys. Letters* 158, 362 (1989)
- [43] D.C. Clary *J. Chem. Phys.* 95, 7298 (1991)
- [44] D. Wang and J. Bowman *J. Chem. Phys.* 96, 8906 (1992)
- [45] J. M. Bowman, *J. Chem. Phys.* **95**, 4960 (1991); J. M. Bowman, *Lecture Notes in Chemistry*, 75, 101 (2000).
- [46] U. Manthe and W. H. Miller, *J. Chem. Phys.* **99**, 3411 (1993).
- [47] D. C. Clary, *J. Phys. Chem.* **98**, 10678 (1994).

- [48] M. Baer and D. J. Kouri, in *The Theory of Chemical Reaction Dynamics*, ed. D.C. Clary (Reidel, Dordrecht, 1986), pp. 167-192.
- [49] G. D. Billing, *J. Chem. Phys.* **107**, 4286 (1997).
- [50] G. D. Billing, *J. Chem. Phys.* **110**, 5526 (1999); *Chem. Phys.* **242**, 341 (1999).
- [51] R. M. Dreizler and E. K. U. Gross, *Density Functional Theory* (Springer Verlag, Berlin, 1990).
- [52] V. Bakken, J. M. Millam, and H. B. Schlegel, *J. Chem. Phys.* **111**, 8773 (1999).
- [53] R. Car, M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
- [54] M. E. Tuckerman, P. J. Ungar, T. von Rosenvinge, and M. L. Klein, *J. Phys. Chem.* **100**, 12878 (1996).
- [55] For a review of *ab initio* classical trajectory methods see K. Balton, W. L. Hase, and G. H. Peshlherbe, in *Modern Methods for Multidimensional Dynamics Computation in Chemistry*, edited by D. L. Thompson (World Scientific, Singapore, 1998), p. 143.
- [56] F. Bernardi, M. Olivucci, M. A. Robb, *J. Am. Chem. Soc.* **114**, 1606 (1992).
- [57] M. Vanneschi, *Lecture Notes in Chemistry* 75, 168 (2000)
- [58] A. Hey, C.J. Scott, M. Surridge, C. Upstill, Integrating computation and information resources, an MPP perspective, in: *Proceedings of 3rd International Working Conference on Massively Parallel Programming Models*, (IEEE Computer Society Press, Los Alamitos, 1998), 174–184.
- [59] M. Vanneschi, PQE2000: HPC tools for industrial applications, *IEEE Concurrency* **6** (1998) 68–73.
- [60] <http://www3.niu.edu/mpi/>
- [61] B. Bacci, B. Cantalupo, P. Pesciullesi, R. Ravazzolo, A. Riaudo, and L. Vanneschi, Skie user guide (version 2.0), Technical report, QSW Ltd., Roma, Italia, December 1998.
- [62] M. Vanneschi, The programming model of Assist, an environment of parallel and distributed portable applications, *Parallel computing* **28**, 12 (2002)
- [63] P. Ciullo, M. Danelutto, S. Magini, L. Potiti, *ASSIST-CL User Manual*, Technical report, ASI-PQE2000, 2002.
- [64] S. Gregori, S. Tasso, A. Laganà, *Lecture Notes in Computer Science*, **3044** 437-444 (2004).
- [65] A. Laganà, ASI-PQE2000 WP4 Technical report 003-005, March 2002.

- [66] Foster I., Kesselman C. Eds., The Grid Blueprint for a Future Computing Infrastructure, Morgan Kaufmann: San Francisco CA, 1999.
- [67] Foster I., Kesselman C. Tuecke S. The anatomy of the grid: Enabling scalable virtual configurations, *Int. J. Supercomp. Appl.* 15 (2001).
- [68] Enabling Grids for E-Science in Europe (EGEE) <http://www.eu-egee.org>
- [69] Laganà, A., Crocchianti, S., Bolloni, A., Piermarini, V., Baraglia, R., Ferrini, R., Laforenza, D.: Computational granularity and parallel models to scale up reactive scattering calculations. *Computer Physics Communications* **128/1-2** 295 (2000)
- [70] First Egee Application Migration Report, EGEE-DNA4.3.1.523422-v4-7-13.doc (2005); Egee review workshop, (<http://agenda.cern.ch/fullAgenda.php?ida=a043803>) Geneva, 2005.
- [71] Preparatory wokshop "METACHEM", Brussels, November, 1999 and Action D23 of COST in Chemistry Memorandum of understanding (see <http://costchemistry.epfl.ch/docs/D23/d23-main.htm>, <http://www.unil.ch/cost/chem>).
- [72] Storchi, L., Tarantelli, F., Laganà, A.: Computing Molecular Energy Surfaces on a Grid, ICCSA 2006, Glasgow (2006).
- [73] Arteconi, L., Pacifici, L., Laganà, A.: Fitting potential energy surfaces for grid enabled molecular simulators, ICCSA 2006, Glasgow (2006).